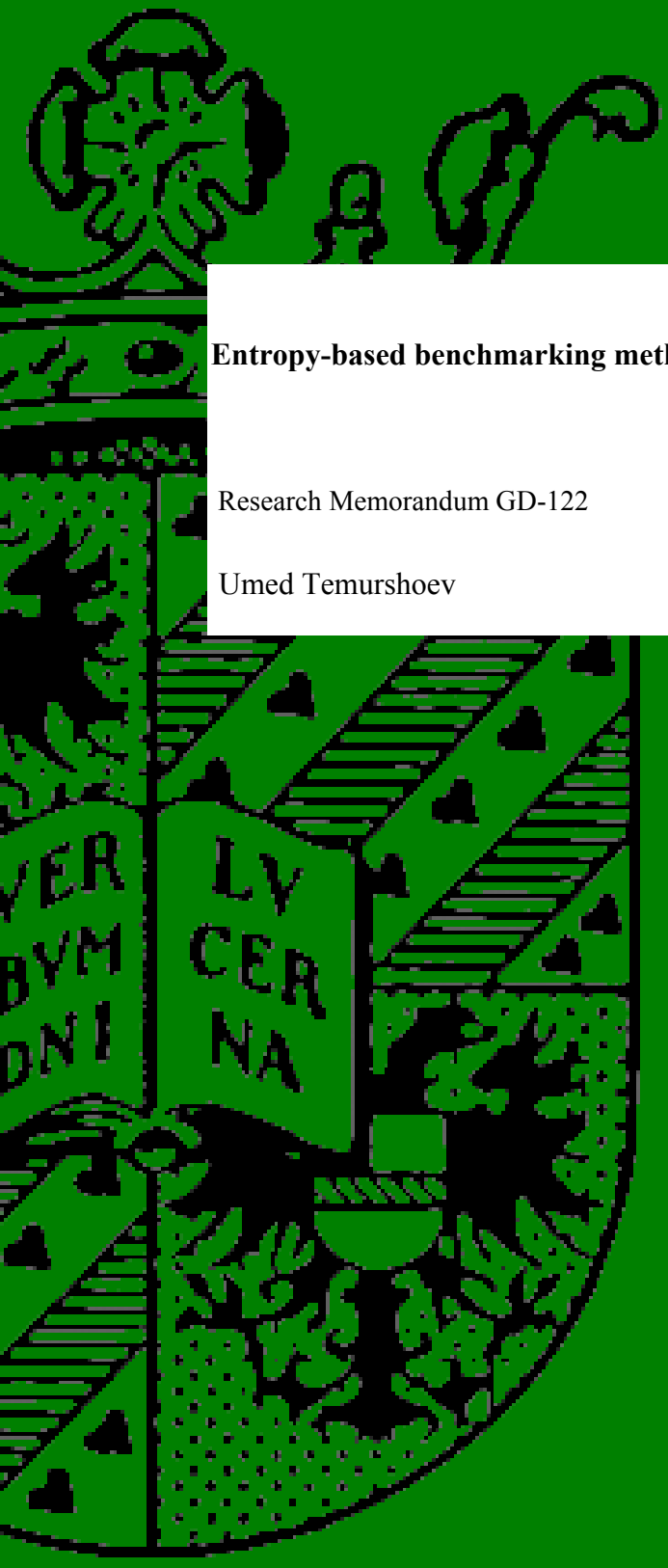
Entropy-based benchmarking methods

Temurshoev, Umed

University of Groningen

Groningen Growth and Development Centre

RuG



Entropy-based benchmarking methods

Research Memorandum GD-122

Umed Temurshoev

RESEARCH MEMORANDUM

Entropy-based benchmarking methods*

Umed Temurshoev†

February 8, 2012

Abstract

We argue that benchmarking sign-volatile series should be based on the principle of movement and sign preservation, which states that a benchmarked series should reproduce the movement and signs in the original series. We show that the widely used variants of Denton (1971) method and the growth preservation method of Causey and Trager (1981) may violate this principle, while its requirements are explicitly taken into account in the proposed entropy-based benchmarking methods. Our illustrative examples show that the entropy-based methods can be regarded as plausible competitors for current benchmarking methods, and maybe preferred in certain cases.

Keywords: benchmarking, generalized cross-entropy estimator, mathematical models

JEL Classification Codes: C02, C61, C80

*The author is grateful to Jacco Daalmans, Tommaso Di Fonzo, Nino Mushkudiani, Reinier Bikker and Jose M. Pavía for their constructive comments and suggestions on the earlier drafts of this paper. Financial support provided by EU FP7 WIOD project is gratefully acknowledged. This project is funded by the European Commission, Research Directorate General as part of the 7th Framework Programme, Theme 8: Socio-Economic Sciences and Humanities. Grant Agreement no: 225 281, www.wiod.org.

†Faculty of Economics and Business, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands. E-mail: u.temurshoev@rug.nl

1 Introduction

Benchmarking problem arises when time series data for the same target variable measured at different time frequencies are inconsistent. That is, benchmarking deals with the problem when a series of high-frequency data (e.g., quarterly data) do not match the same series of low-frequency data (e.g., annual data). The low-frequency data, called in the literature as the *benchmark series*, are usually of high precision and provide reliable information on the aggregate level and long-term movements of economic time series. On the other hand, low-frequency data are less precise but are timely and provide the only information about the short-term movements in the series (see e.g. Dagum and Cholette 2006). Hence, the problem is to estimate/revise the available high-frequency data, also called the *preliminary* or *indicator series*, such that they are in line with the fixed benchmark series. Such adjustments, however, have to maximally take the relative strengths of the benchmark and preliminary series into account. That is, the obtained high-frequency data should preserve as much as possible the short-term movements in the indicator series under the restrictions stemming from the benchmark series that in turn display the long-term movements of the same data.[1]

Two types of methods have been developed for benchmarking problems: mathematical and statistical methods. The main distinction between these approaches is that a mathematical (resp. statistical) model treats the estimation process of the unknown high-frequency series as deterministic (resp. stochastic) and considers the benchmark (annual) constraints as binding (resp. either binding or non-binding). The most widely used mathematical approaches are the variants of Denton (1971) adjustment method, and Causey and Trager (1981) growth rate preservation method. The most commonly used statistical approaches are the variants of Chow and Lin (1971) regression methods, and time-series ARIMA and generalized regression-based methods (for details see e.g., Dagum and Cholette 2006).

Chen (2007) is an excellent recent study that assesses the performance of various mathematical and statistical methods for benchmarking using 60 series from the US national economic accounts. The conclusion of this work is that "the modified Denton proportional first difference method outperforms the other methods,

[1]The related problem, which is not the focus of this paper, is the so-called *temporal disaggregation*. It deals with the similar problem where a preliminary series is *not* for the same target variable as the benchmark series. For example, total taxes can be used as the preliminary series for taxes on wholesale trade. Another important difference is that while benchmarking involves one preliminary series vis-a-vis one benchmark series, in temporal disaggregation one might have several indicator series (often measured in different units).

though the Causey-Trager growth preservation model is a close competitor in certain cases". Further, Chen and Andrews (2008) state that in the US "[a]ccording to the current revision policy, quarterly or monthly estimates of the 3 most recent years are revised during annual revisions" (p. 32). Evidently, such revision policies do not allow statistical agencies to use time series techniques because the number of yearly observations is quite small. Given this restriction and the results of Chen's (2007) study, the "Bureau of Economic Analysis has adopted a new method for the interpolation of quarterly and monthly estimates in the national accounts. The new method uses a variant of the Denton procedure" (Chen and Andrews 2008, p. 31). In fact, the variants of Denton's procedure are widely used by national statistical institutes of other countries as well (see e.g., Bikker et al. 2010). On the other hand, the growth preservation method of Causey and Trager (1981) is used by the US Census Bureau for benchmarking monthly and quarterly series to the corresponding annual survey values (see e.g., Brown 2010, Titova et al. 2010). Thus, in this study we also focus on the above mentioned mathematical methods.

The purpose of this paper is twofold. First, we argue that, in general, benchmarking methods should be based on the *principle of movement and sign preservation*, which states that (i) short-term movements in the observed high-frequency data should be reproduced in its benchmarked series and (ii) the signs of the original elements should be kept in the benchmarked series. In the literature, however, much attention is given only to the principle of *movement preservation* which has to do with only the first part of the movement and sign preservation principle. That is, it states that the benchmarked series should reproduce the movement in the original series. The justification is that the short-term movement in the original data is the only available information. However, we think once the series with both positive and negative elements are considered, this principle becomes insufficient because extra important information about the signs of the original elements is disregarded and the movement preservation does not guarantee signs preservation. Given that there are abundant socio-economic series that are volatile and allow for both positive and negative values, we believe that instead the principle of movement and sign preservation should be the basis for benchmarking methods.

Second, we propose two alternative benchmarking methods that by construction take care of the mentioned principle. These methods are based on entropy formalism, which takes its origin from information theory as developed by Shannon (1948) and the work of Kullback and Leibler (1951), Janes (1957a) and Janes (1957b). We write the additive first difference (AFD) and proportional first difference (PFD)

benchmarking principles in a form consistent with a generalized cross-entropy (GCE) framework of Golan et al. (1996). The important features of the AFD and PFD entropy-based benchmarking methods are the possibility of using any prior beliefs or non-sample information about the signs, magnitudes, and/or ranges of plausible values of any time series within the GCE framework, flexibility of choosing any kind of binding and/or nonbinding benchmarking constraints, their applicability to any number of yearly observations for revision purposes, and access to the derived reliability estimates (factors) of each element of the benchmarked series.

We also provide theoretical foundation for the so-called pro rata distribution method and generalize it to benchmarking of time series with both positive and negative elements. In our illustrative examples we perform an assessment test of the generalized pro rata distribution, the widely used AFD and PFD variants of the Denton procedure, the Causey-Trager growth rate preservation method, and the proposed entropy-based benchmarking methods. The results show that the entropy-based methods do perform as well as the widely used benchmarking methods, and maybe even preferred in cases of sign-volatile series benchmarking.

The rest of the paper is organized as follows. In Section 2 we briefly discuss the variants of Denton method and the Causey-Trager growth preservation model and show why they may violate the movement and sign preservation principle. In Section 3 we generalize the pro rata distribution to benchmarking of sign-volatile time series, thus provide theoretical foundation to this method. In Section 4 we present the entropy-based benchmarking methods. All the methods are compared on two illustrative examples of benchmarking in Section 4. Section 5 concludes.

2 Denton and Causey-Trager methods

In benchmarking problems we have a preliminary (or indicator) series that do not satisfy its benchmark (e.g., annual) constraints. Hence, the problem is estimating a new high-frequency series as close as possible to the indicator series that are consistent with the low-frequency (benchmark) series of the same target variable.

2.1 Preliminaries

Without loss of generality, assume that the time series of interest covers $t = 1, \ldots, T$ years and each year series has $m = 1, \ldots, M$ sub-periods. Let $\mathbf{z}_t$ and $\mathbf{x}_t$ be, respectively, the $M \times 1$ vectors of observed (indicator) and estimated high-frequency

(benchmarked) series for year t.[2] Thus, the preliminary and benchmarked series can be compactly written by the $MT \times 1$ vectors $\mathbf{z} = (\mathbf{z}_1', \ldots, \mathbf{z}_T')'$ and $\mathbf{x} = (\mathbf{x}_1', \ldots, \mathbf{x}_T')'$. Further, the benchmark (say, annual) series is denoted by the $T \times 1$ vector $\mathbf{y} = (y_1, \ldots, y_T)'$.[3] For example, y_t can be the annual total of a sub-annual series of year t for some flow variable (say, GDP), or the value of the last element of a sub-annual series of year t for certain stock variable (say, population). Let the $M \times 1$ vector $\mathbf{b}$ be a *generalized* coefficients vector that defines the benchmarking constraints of interest. For the above mentioned cases, for example, we will have $\mathbf{b} = (1, \ldots, 1)' = \boldsymbol{\imath}$ for flow variables and $\mathbf{b} = (\mathbf{0}', 1)'$ for stock variables. Consider the following $T \times MT$ matrix

$$\mathbf{B} = \begin{bmatrix} \mathbf{b}' & \mathbf{0}' & \vdots & \mathbf{0}' \\ \mathbf{0}' & \mathbf{b}' & \vdots & \mathbf{0}' \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}' & \mathbf{0}' & \cdots & \mathbf{b}' \end{bmatrix},$$

which can simply be defined as $\mathbf{B} = \mathbf{I}_T \otimes \mathbf{b}'$ with $\mathbf{I}_T$ and $\otimes$ denoting the T-square identity matrix and the Kronecker product, respectively. Hence, the benchmark-consistency constraints can be compactly written as

$$\mathbf{B}\mathbf{x} = \mathbf{y}. \tag{1}$$

Obviously, the benchmarking problem arises because $\mathbf{B}\mathbf{z} \neq \mathbf{y}$, i.e., the original series $\mathbf{z}$ is not consistent with the benchmark series $\mathbf{y}$. Thus, there is a need for adjustments in $\mathbf{z}$ that results in a new high-frequency data $\mathbf{x}$ such that (1) holds.[4]

2.2 Denton's benchmarking methods

Denton (1971) used the following penalty function as a measure of the distance between the indicator and benchmarked series:

[2]Matrices are given in bold capital letters; vectors in bold lower case letters; and scalars in italicized lower/capital case letters. Vectors are columns by definition, thus row vectors are obtained by transposition, indicated by a prime. The null vector and the summation vector of ones of appropriate dimensions are denoted by $\mathbf{0}$ and $\boldsymbol{\imath}$, respectively.

[3]Note the difference between the benchmark and benchmarked series. While the benchmark series represents the fixed low-frequency data, the benchmarked series is the adjusted high-frequency data that is consistent with the benchmarking constraints.

[4]In fact, the inequality in $\mathbf{B}\mathbf{z} \neq \mathbf{y}$ does *not* have to be element-wise inequality and generally means the existence of at least one strict inequality sign. That is, there might exist such $\mathbf{z}_k$'s not for all years that are consistent with their benchmark values, i.e., $\mathbf{b}'\mathbf{z}_k = y_k$.

$$f(\mathbf{x}, \mathbf{z}, \mathbf{A}) = (\mathbf{x} - \mathbf{z})' \mathbf{A} (\mathbf{x} - \mathbf{z}). \tag{2}$$

The form of the matrix $\mathbf{A}$ defines the following variants of the *original* Denton method for benchmarking:

1. *additive level difference* (ALD) variant if $\mathbf{A} = \mathbf{I}$,
2. *additive first difference* (AFD) variant if $\mathbf{A} = \mathbf{D}'\mathbf{D}$,
3. *additive second difference* (ASD) variant if $\mathbf{A} = (\mathbf{D}^2)'\mathbf{D}^2$,
4. *proportional level difference* (PLD) variant if $\mathbf{A} = \hat{\mathbf{z}}^{-1}\hat{\mathbf{z}}^{-1}$,
5. *proportional first difference* (PFD) variant if $\mathbf{A} = \hat{\mathbf{z}}^{-1}\mathbf{D}'\mathbf{D}\hat{\mathbf{z}}^{-1}$, and
6. *proportional second difference* (PSD) variant if $\mathbf{A} = \hat{\mathbf{z}}^{-1}(\mathbf{D}^2)'\mathbf{D}^2\hat{\mathbf{z}}^{-1}$,

where

$$\underset{T\times T}{\mathbf{D}} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}. \tag{3}$$

In what follows we will discuss only the AFD and PFD variants of the Denton procedure since namely these methods are widely used in practice. It is not difficult to derive that the AFD and PFD penalty functions have the following forms:

$$f_{AFD} = (x_1 - z_1)^2 + \sum_{j=2}^{N} \left[(x_j - z_j) - (x_{j-1} - z_{j-1}) \right]^2, \tag{4}$$

$$f_{PFD} = \left(\frac{x_1}{z_1} - 1 \right)^2 + \sum_{j=2}^{N} \left(\frac{x_j}{z_j} - \frac{x_{j-1}}{z_{j-1}} \right)^2, \tag{5}$$

where $N = MT$. From (4) it follows that with the AFD Denton procedure the benchmarking problem boils down to minimization of the overall sum of squares of the *period-to-period level changes* of the benchmarked and indicator series subject to benchmarking constraints (1). The PFD Denton approach (5), on the other hand, is based on the minimization of the sum of squares of the corresponding *period-to-period proportional changes*, again subject to the benchmarking constraints (1).

Note that Denton imposed the following initial conditions: $x_0 = z_0$ for his AFD method, and $x_0 = z_0$ and $x_{-1} = z_{-1}$ for his PFD approach. These conditions essentially mean that no adjustments to the original series are allowed outside the range of a sample. It has been argued that these initial conditions introduce a

transient movement at the beginning of the series that goes at odd with the principle of movement preservation (see e.g., Helfand et al. 1977, Dagum and Cholette 2006). This issue is solved in the so-called *modified* Denton procedure, where the first terms in (4) and (5) are omitted. The list of the variants of the modified Denton method are similar to the list provided above with minor changes: for the AFD and PFD variants instead of $\mathbf{D}$ in (3) one uses the matrix $\mathbf{D}_f$, whereas for the ASD and PSD variants instead of $\mathbf{D}^2$ one uses the matrix $\mathbf{D}_s$ that are defined as follows:[5]

$$\underset{(T-1)\times T}{\mathbf{D}_f} = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}, \quad \underset{(T-2)\times T}{\mathbf{D}_s} = \begin{bmatrix} 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & -2 & 1 \end{bmatrix}. \quad (6)$$

It is important to find out what are in fact the benchmarked series x_j equal or close to when the widely used AFD and PFD Denton methods are applied in practice. Let us for the moment consider a case when the benchmarking constraints (1) are ignored. Then the optimal condition for the AFD Denton problem, $\partial f_{AFD}/\partial x_j = 0$, results in the following solution

$$x_j^* = z_j + \frac{(x_{j+1}^* - z_{j+1}) + (x_{j-1}^* - z_{j-1})}{2}, \qquad j \neq 1, N, \quad (7)$$

i.e., the difference between the benchmarked and indicator values, $x_j^* - z_j$, is exactly equal to the average of the two neighboring corresponding level differences. If the benchmarking constraints (1) are not satisfied, then the problem must take them into account (in a Lagrangian setting). However, in that case also depending on the degree of violation of the benchmark-consistency constraints the solution will *tend to* be as close as possible to the solution (7).

In the case of the PFD Denton method the first-order condition, $\partial f_{PFD}/\partial x_j = 0$, again ignoring (1), yields

$$x_j^* = \frac{z_j}{2}\left(\frac{x_{j+1}^*}{z_{j+1}} + \frac{x_{j-1}^*}{z_{j-1}}\right), \qquad j \neq 1, N, \quad (8)$$

i.e., the ratio of the benchmarked and indicator values, x_j^*/z_j, is exactly equal to the arithmetic average of the two neighboring corresponding ratios. Whenever the benchmarking constraints (1) are taken care of, the solution of the PFD Denton

[5]The solution of all variants of Denton procedure are given in equation (2.2) in Denton (1971, p. 100). We should note that in terms of notations Denton's $\mathbf{B}$ is our $\mathbf{B}'$.

method again will *tend to* be as close as possible to that given in (8).

These observations imply the following important outcomes.

1. The AFD and PFD variants of the Denton method are, in fact, *smoothing* methods. For example, from (7) it follows that if $z_j = 0$ but its adjacent level differences $(x^*_{j-1} - z_{j-1})$ and $(x^*_{j+1} - z_{j+1})$ are positive, then the estimated x_j will be positive either (and not zero).[6] In the literature, however, it is usually claimed that only the first- and second-difference variants of Boot et al. (1967) approach are smoothing methods, which are used for generating high-frequency data in the absence of benchmark series (see e.g., Eurostat 1999, p. 155).
2. The first observation also implies that the proportional difference versions of the Denton procedure better preserve (close to) zero values from $\mathbf{z}$ onto $\mathbf{x}$ than their corresponding additive difference variants (see e.g., (8)). This partly explains why the PFD Denton method is in many cases preferred by benchmarking practitioners over its AFD version.
3. If the indicator series are volatile and admit both positive and negative values, then, provided that the benchmarking constraints are taken into account, (7) and (8) imply the possibility of obtaining an unacceptable sign-changing outcome. That is, one can have $x^*_j < 0$ with $z_j > 0$ and vice versa. Thus, the AFD and PFD Denton methods violate the principle of movement and sign preservation.

2.3 Causey and Trager's benchmarking method

Causey and Trager (1981) consider a different non-linear minimization problem, where the target of preservation is the *period-to-period growth rate* of the indicator series. The corresponding objective function has the form

$$f_{CT} = \sum_{j=2}^{N} \left(\frac{x_j}{x_{j-1}} - \frac{z_j}{z_{j-1}} \right)^2 . \tag{9}$$

Causey-Trager (CT) growth rate preservation method is referred to as the "ideal" benchmarking method because growth rate is considered to be a natural measure of movement of time series (see e.g., Helfand et al. 1977, Di Fonzo and Marini 2010).

[6] In empirical applications of the modified PFD Denton method, whenever $z_j = 0$ its inverse $1/z_j$ is not defined. If such cases arise, one may simply replace them with small numbers. It should be mentioned that the original Denton method avoids using these reciprocals, hence there is no need for bothering about zero entries in $\mathbf{z}$ (see equation (3.4) and discussions of its practical use in Denton (1971), p. 100).

In contrast to Denton method, the CT method that is based on minimization of (9) subject to (1) does *not* have analytical solution. Hence, numerical methods are called for in order to find the solution of the CT benchmarking approach.[7]

As before, if we ignore the benchmark-consistency constraints (1), the first-order condition of the CT problem, $\partial f_{CT}/\partial x_j = 0$, can be derived as

$$x_j^* = x_{j-1}^* \left[\frac{z_j}{z_{j-1}} + \left(\frac{x_{j+1}^*}{x_j^*} - \frac{z_{j+1}}{z_j} \right) \frac{x_{j+1}^*/x_j^*}{x_j^*/x_{j-1}^*} \right], \qquad j \neq 1, N. \tag{10}$$

Equation (10) implies that ignoring benchmarking constraints, the optimal growth rates in the benchmarked series, x_j^*/x_{j-1}^*, equal the corresponding growth rates of the indicator series, z_j/z_{j-1}, plus the difference of the next period growth rates of the benchmarked and indicator series, weighted by the ratio of the benchmarked growth rates of the next and current periods. When benchmarking constraints are taken into account, the CT outcome *tends to* be as close as possible to the solution (10). Thus, all three observations made earlier with respect to the AFD and/or PFD variants of the Denton method also hold for the CT solution.

3 Generalized pro rata distribution method

Consider the case when the preliminary and benchmark series include only positive elements. In this section we focus on benchmarking constraints as the sum of flow variables, i.e., $\mathbf{b} = \boldsymbol{\imath}$.[8] The pro rata distribution makes the indicator series consistent with its benchmark series simply by distributing the benchmark values (annual totals) in proportion to the shares of the sub-period elements in the corresponding indicator aggregates, i.e.,

$$\mathbf{x}_t = \mathbf{z}_t \times \left(\frac{y_t}{\boldsymbol{\imath}'\mathbf{z}_t} \right) \qquad \text{for all } t = 1, \ldots, T. \tag{11}$$

From (11) it is evident that adjustments of the sub-period elements in one year of the indicator series is completely independent from that of the other years. Thus,

[7] For finding the CT solution, following Di Fonzo and Marini (2010) we use MATLAB built-in function *fmincon* and its option of *interior-point* algorithm.

[8] A more general form for flow variables is $\mathbf{b} = a \times \boldsymbol{\imath}$, where a is a nonnegative real number. For example, if the annual constraints represent the averages of quarterly data, then $a = 1/4$. However, all these forms can be easily transformed into the simple summation conditions with unit weights by dividing the observed data by a. Then the derived solution has to be translated back to the original scales by multiplying by a. Pro rata distribution approach to benchmarking stock variables is discussed in the Appendix.

pro-rata distribution method introduces unacceptable discontinuities into the benchmarked time series, which can be easily recognized by an abrupt upward and/or downward steps in the charts of the benchmark-to-indicator ratios. In the benchmarking literature this is known as the *step problem*, which evidently goes at odd with the principle of movement preservation.

How does the pro rata distribution work when the time series of interest includes both positive and negative elements? It turns out that this method also has a theoretical foundation similar to the Denton and Causey-Trager approaches discussed in the previous section. Consider the following objective function:

$$f_{GRAS} = \sum_{j=1}^{N} |z_j| \left(h_j \ln \left(\frac{h_j}{e} \right) + 1 \right), \tag{12}$$

where $|z_j|$ is the absolute value of the indicator series z_j, e is the base of the natural logarithm, and the benchmark-to-indicator ratio is defined as $h_j \equiv x_j/z_j$ for all $j = 1, \ldots, N$ whenever $z_j \neq 0$ and $h_j = 1$ for $z_j = 0$.

Function (12) is the objective used in the so-called generalized RAS (GRAS) framework for updating input-output tables by Junius and Oosterhaven (2003) and Lenzen et al. (2007), and for estimating supply and use tables by Temurshoev and Timmer (2011). The GRAS function is also somewhat similar to the well-known information-based entropy measure. Minimizing this function subject to certain constraints implies that the benchmarked value x_j is *as close as possible* to the corresponding indicator value z_j for all j. To see this, first, observe that for $h_j = 1$ the value of (12) is zero, which is its minimum possible value. Second, if we ignore the benchmarking constraints, the first-order condition, $\partial f_{GRAS}/\partial h_j = 0$, gives $\ln(h_j) = 0$, which is equivalent to

$$x_j^* = z_j. \tag{13}$$

Whenever the GRAS function (12) is minimized subject to the benchmarking constraints $\boldsymbol{\iota}'\mathbf{x}_t = y_t$ for all $t = 1, \ldots, T$, the solution will *tend to* be as close as possible to that given in (13). Thus, compared to the AFD and PFD Denton procedures and the Causey-Trager (CT) growth preservation method, the GRAS-like benchmarking approach will ensure that the *level* of each element in the benchmarked series is maximally close to the corresponding level of the indicator series (compare (13) to (7), (8) and (10)).

Let us define $\mathbf{p}_t$ as a vector with all non-negative entries of $\mathbf{z}_t$ and $\mathbf{n}_t$ as a vector

containing absolute values of the negative elements of $\mathbf{z}_t$, i.e., $\mathbf{z}_t = \mathbf{p}_t - \mathbf{n}_t$ for all $t = 1, \ldots, T$. In the Appendix we show that the solution of minimization of (12) subject to the benchmarking constraints $\boldsymbol{\imath}'\mathbf{x}_t = y_t$ is

$$\mathbf{x}_t = \mathbf{p}_t \times s_t - \mathbf{n}_t / s_t \qquad \text{for all } t = 1, \ldots, T, \tag{14}$$

where $s_t > 0$ is an *annual adjustment factor* for year t and is defined as

$$s_t = \begin{cases} 0.5 \times \left[y_t + \sqrt{y_t^2 + 4 \times (\boldsymbol{\imath}'\mathbf{p}_t) \times (\boldsymbol{\imath}'\mathbf{n}_t)} \right] / \boldsymbol{\imath}'\mathbf{p}_t & \text{if } \boldsymbol{\imath}'\mathbf{p}_t > 0, \\ -\boldsymbol{\imath}'\mathbf{n}_t / y_t = \boldsymbol{\imath}'\mathbf{z}_t / y_t & \text{if } \boldsymbol{\imath}'\mathbf{p}_t = 0. \end{cases} \tag{15}$$

Consider a time series of positive elements only. Then, we have that $\mathbf{p}_t = \mathbf{z}_t$ and $\mathbf{n}_t = \mathbf{0}$, and therefore (14) and (15) together boil down to the standard pro rata distribution formula (11). If, on the other hand, the time series of interest consists of only non-positive elements, $\mathbf{z}_t < \mathbf{0}$, then $\mathbf{p}_t = \mathbf{0}$ and $\mathbf{n}_t = -\mathbf{z}_t$. Thus, from (15) it follows that $s_t = \boldsymbol{\imath}'\mathbf{z}_t / y_t$, which is still positive as in this case it must be also true that $y_t < 0$. Thus, (14) boils down to $\mathbf{x}_t = -\mathbf{n}_t / s_t = \mathbf{z}_t y_t / (\boldsymbol{\imath}'\mathbf{z}_t)$, which is again the standard pro rata distribution solution. Therefore, we call the solution (14)-(15) as a *generalized* pro rata distribution method since it allows the indicator series to include both positive and negative elements (in an arbitrary order).

In the generalized pro rata distribution (14) all the positive elements of the indicator series of year t are adjusted by the same adjustment factor $s_t > 0$, while all the strictly negative entries of $\mathbf{z}_t$ are adjusted by $1/s_t > 0$. The intuition for this proportional adjustment is simple. Assume that $\boldsymbol{\imath}'\mathbf{z}_t < y_t$, then in order to ensure the benchmarking constraint $\boldsymbol{\imath}'\mathbf{x}_t = y_t$ it must be the case that during the adjustment procedure all the positive entries in the indicator series for year t increase, while the absolute values of all the negative elements in $\mathbf{z}_t$ decrease. This is exactly what happens with (14) with $s_t > 1$. Since the annual adjustment factor is a strictly positive number, from (14) also follows that the signs of the elements of the indicator series are kept into the benchmarked series. The sign preservation property is another distinguishing feature of the generalized pro rata distribution from the AFD and PFD Denton methods and CT benchmarking approach. However, as we noted above the generalized pro rata distribution violates the principle of movement preservation due to the resulting step problem.

4 Entropy-based benchmarking methods

In this section we first discuss briefly a generalized cross-entropy approach to recovering parameters in underdermined or ill-posed systems. Then this approach is applied to benchmarking problems when the principle of movement preservation is defined in terms of additive first difference and proportional first difference preservation of the indicator series.

4.1 Generalized cross-entropy approach

We consider the following linear inverse problem with noise formulation (see Golan et al. 1996):

$$\tilde{\mathbf{y}} = \mathbf{\Gamma}\mathbf{x} + \mathbf{e}, \tag{16}$$

where $\tilde{\mathbf{y}}$ is an I-dimensional vector of observables (not to be confused with the benchmark series $\mathbf{y}$), $\mathbf{x}$ is a N-dimensional vector of unknown parameters (or the benchmarked series for our purposes), $\mathbf{\Gamma}$ is a known non-square linear operator matrix, and $\mathbf{e}$ is a disturbance vector. It is often the case that $N > I$ implying that the system (16) is undertermined or ill-posed. That is, in a familiar regression setting we will have the number of parameters larger than the number of observations, thus the traditional methods such as OLS cannot be applied to estimate $\mathbf{x}$.

Entropy-based methods can handle ill-posed problems, which are the core of a generalized cross-entropy (GCE) approach developed by Golan et al. (1996). Entropy formalism is based on information theory as developed by Shannon (1948) and the studies of Kullback and Leibler (1951), Janes (1957a) and Janes (1957b).

The GCE starting point is that one has to use his/her limited prior or non-sample information about the unknown parameters $\mathbf{x}$ (also called the signal component) and noise component $\mathbf{e}$. This information includes prior beliefs about the signs, magnitudes, and/or ranges of plausible values of these unknown components. Then one constructs discrete random variables with prior weights (probabilities) and finite supports that is consistent with the given non-sample information about $\mathbf{x}$ and $\mathbf{e}$.[9] Accordingly, the linear inverse problem is reparametrized in terms of discrete random variables on bounded supports, and the estimation problem becomes recovering "posterior" probability distributions for $\mathbf{x}$ and $\mathbf{e}$ consistent with the available prior information and the observed sample information.

[9] "It is important to note that the random variables are merely conceptual devices used to express the prior and sample knowledge in a mutually compatible format" (Golan et al. 1996, p. 86).

Following Golan et al. (1996), let us first represent $\mathbf{x}$ by expectations of random variables with compact supports. That is, x_j is treated as a discrete random variable with a compact support of K possible outcomes $\mathbf{r}_j = (r_{j1}, \ldots, r_{jK})'$, where $2 \leq K < \infty$, and r_{j1} and r_{jK} are the plausible lower and upper bounds (extreme values) of x_j. Hence, x_j is expressed as a convex combination

$$x_j = \sum_{k=1}^{K} r_{jk}\tilde{p}_{jk} = \mathbf{r}_j'\tilde{\mathbf{p}}_j,$$

where $\tilde{\mathbf{p}}_j$ is a K-dimensional vector of positive weights (probabilities) that sum to one (not to be confused with $\mathbf{p}_t$ from Section 3). These convex combinations can be assembled in matrix form so that $\mathbf{x}$ may be written as

$$\mathbf{x} = \mathbf{R}\tilde{\mathbf{p}} = \begin{bmatrix} \mathbf{r}_1' & \mathbf{0}' & \cdots & \mathbf{0}' \\ \mathbf{0}' & \mathbf{r}_2' & \cdots & \mathbf{0}' \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}' & \mathbf{0}' & \cdots & \mathbf{r}_N' \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{p}}_1 \\ \tilde{\mathbf{p}}_2 \\ \vdots \\ \tilde{\mathbf{p}}_N \end{bmatrix}, \tag{17}$$

where $\mathbf{R}$ is a $N \times NK$ matrix, $\tilde{\mathbf{p}}$ is a NK-dimensional vector of weights, and $\mathbf{0}$ is a null vector of appropriate dimension (here it is a K-dimensional vector).

Similarly, one can represent his/her uncertainty about the outcome of the error process by treating each e_i as a finite discrete random variable with $2 \leq J < \infty$ possible outcomes. Suppose there exist extreme values v_{i1} and v_{iJ} for each e_i such that $1 - Pr(v_{i1} < e_i < v_{iJ})$ may be made arbitrarily small. Each disturbance then can be written as

$$e_i = \sum_{j=1}^{J} v_{ij}w_{ij} = \mathbf{v}_i'\mathbf{w}_i,$$

where $\mathbf{v}_i = (v_{i1}, \ldots, v_{iJ})'$ is a finite support for e_i and $\mathbf{w}_i = (w_{i1}, \ldots, w_{iJ})'$ is a J-dimensional vector of positive weights (probabilities) that sum to one. Thus, the I-dimensional vector of unknown disturbances can be compactly written as

$$\mathbf{e} = \mathbf{V}\mathbf{w} = \begin{bmatrix} \mathbf{v}_1' & \mathbf{0}' & \cdots & \mathbf{0}' \\ \mathbf{0}' & \mathbf{v}_2' & \cdots & \mathbf{0}' \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}' & \mathbf{0}' & \cdots & \mathbf{v}_I' \end{bmatrix} \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \vdots \\ \mathbf{w}_I \end{bmatrix}, \tag{18}$$

where $\mathbf{V}$ is an $I \times IJ$ matrix and $\mathbf{w}$ is an IJ-dimensional vector of positive weights.

Using the reparametrized unknowns (17) and (18), model (16) can be rewritten as

$$\tilde{\mathbf{y}} = \mathbf{\Gamma x} + \mathbf{e} = \mathbf{\Gamma R}\tilde{\mathbf{p}} + \mathbf{Vw}. \tag{19}$$

The non-sample information about $\mathbf{x}$ and $\mathbf{e}$ may be expressed as a set of subjective probability distributions on their corresponding supports $\mathbf{R}$ and $\mathbf{V}$. Let $\mathbf{q}$ be the NK dimensional vector of *prior* weights for the unknown parameters $\mathbf{x}$, hence the *prior mean* for $\mathbf{x}$ is $\mathbf{Rq}$. Similarly, let $\mathbf{u}$ be the IJ-dimensional vector of prior weights on disturbances $\mathbf{e}$ with prior mean $\mathbf{Vu}$. The GCE estimator is defined by

$$\min_{\tilde{\mathbf{p}},\mathbf{w}} \; I(\tilde{\mathbf{p}}, \mathbf{q}, \mathbf{w}, \mathbf{u}) = \tilde{\mathbf{p}}' \log(\tilde{\mathbf{p}}/\mathbf{q}) + \mathbf{w}' \log(\mathbf{w}/\mathbf{u}) \tag{20}$$

subject to

$$\tilde{\mathbf{y}} = \mathbf{\Gamma R}\tilde{\mathbf{p}} + \mathbf{Vw}, \tag{21}$$

$$\boldsymbol{\imath}_N = (\mathbf{I}_N \otimes \boldsymbol{\imath}_K')\tilde{\mathbf{p}}, \tag{22}$$

$$\boldsymbol{\imath}_I = (\mathbf{I}_I \otimes \boldsymbol{\imath}_J')\mathbf{w}, \tag{23}$$

where / is an element-wise division and the subscripts of the identity matrices and summation vectors indicate their corresponding dimensions.

The *principle of minimum discrimination information* (MDI) proposed by Kullback (1959, pp. 36-43) underlies the GCE objective function (20). It states that the new distributions $\tilde{\mathbf{p}}$ and $\mathbf{w}$, given the model restrictions (21), are chosen (estimated) such that they are minimally discriminated from the original distributions $\mathbf{q}$ and $\mathbf{u}$, respectively. MDI is also often called the principle of minimum cross-entropy. If it turns out that the prior information, $\mathbf{q}$ and $\mathbf{u}$, is consistent with the data, then the MDI solution is $\tilde{\mathbf{p}} = \mathbf{q}$ and $\mathbf{w} = \mathbf{u}$ with $I(\tilde{\mathbf{p}}, \mathbf{q}, \mathbf{w}, \mathbf{u}) = 0$. This would mean that the data has no additional information relative to the prior.

Nonexistence of non-sample information often makes analysts to choose prior as a uniform distribution (consistent with the famous Laplace's principle of insufficient reason). In such a case, the cross-entropy becomes equivalent to a maximum entropy approach. Hence, the MDI can be seen as an extension of the *principe of maximum entropy* of Janes (1957a,b) which states that "in making inferences on the basis of partial information we must use that probability distribution which has maximum entropy subject to whatever is known" (Janes 1957a, p. 623). For further details, the reader is referred to the relevant literature, e.g., Cover and Thomas (2006).

Note that (21) is the model or consistency constraint, while equations (22) and (23) are the required additivity (or normalization) constraints for $\tilde{\mathbf{p}}$ and $\mathbf{w}$, respec-

tively. The GCE objective $I(\tilde{\mathbf{p}}, \mathbf{q}, \mathbf{w}, \mathbf{u})$ is strictly convex on the interior of the additivity constraint set, and if the intersection of the consistency and additivity constraint set is non-empty, then a *unique* solution exists. After setting the Lagrangean it can be shown that the solution of the GCE problem is given by

$$\hat{p}_{nk} = \frac{q_{nk} \exp(r_{nk} \boldsymbol{\Gamma}'_n \hat{\boldsymbol{\lambda}})}{\Omega_n(\hat{\boldsymbol{\lambda}})} \quad \text{and} \quad \hat{w}_{ij} = \frac{u_{ij} \exp(v_{ij} \hat{\lambda}_i)}{\Psi_i(\hat{\lambda}_i)}, \tag{24}$$

where $\boldsymbol{\Gamma}_n$ is the nth column of $\boldsymbol{\Gamma}$, $\boldsymbol{\lambda}$ is an $I \times 1$ vector of Lagrange multipliers, and the normalization factors are

$$\Omega_n(\hat{\boldsymbol{\lambda}}) = \sum_{k=1}^{K} q_{nk} \exp(r_{nk} \boldsymbol{\Gamma}'_n \hat{\boldsymbol{\lambda}}) \quad \text{and} \quad \Psi_i(\hat{\lambda}_i) = \sum_{j=1}^{J} u_{ij} \exp(v_{ij} \hat{\lambda}_i).$$

The GCE solutions, $\hat{\mathbf{p}}$ and $\hat{\mathbf{w}}$, satisfy the additivity constraints and are positive. But since they depend of $\boldsymbol{\lambda}$ that cannot be determined by the problem first-order conditions, there is no closed-form solution. Hence, it must be found numerically and often an efficient computing algorithm is used based on an unconstrained dual GCE formulation (for details see Golan et al. 1996, pp. 93-96).[10] Finally, the optimal probability vectors are used to compute the point estimates of the unknown parameter and disturbance vectors, i.e., $\hat{\mathbf{x}} = \mathbf{R}\hat{\mathbf{p}}$ and $\hat{\mathbf{e}} = \mathbf{V}\hat{\mathbf{w}}$.

4.2 AFD entropy-based benchmarking method

Consider a type of movement preservation based on the additive first difference (AFD) preservation principle. Recall that this principle is the core of the AFD Denton procedure, where the objective is minimization of the sum of squares of period-to-period level changes of the benchmarked and indicator series (see (4)). Minimization of the AFDs, however, can also be approached from a somewhat different perspective. The AFD principle can be alternatively implemented via

$$x_j - z_j = (x_{j-1} - z_{j-1}) - \varepsilon_j \quad \text{for all } j = 2, \ldots, N \tag{25}$$

with as small as possible error term ε_j ($\gtreqless 0$). Define

[10] We use this more efficient computational approach in our empirical applications, where we also employ the analytical gradient and Hessian matrix of the dual problem. The required computations are implemented in MATLAB with *fminunc* function.

$$\Delta\mathbf{z} = \begin{bmatrix} z_2 - z_1 \\ z_3 - z_2 \\ \vdots \\ z_N - z_{N-1} \end{bmatrix}, \quad \mathbf{C}_A = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 & 1 \end{bmatrix}, \quad \text{and} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_N \end{bmatrix}.$$

Since (25) is equivalent to $z_j - z_{j-1} = x_j - x_{j-1} + \varepsilon_j$, in matrix form these AFD constraints can be compactly written as

$$\Delta\mathbf{z} = \mathbf{C}_A\mathbf{x} + \boldsymbol{\varepsilon}. \tag{26}$$

Next, instead of (1), we consider the following benchmarking constraints with non-binding possibilities

$$\mathbf{y} = \mathbf{B}\mathbf{x} + \boldsymbol{\tau}, \tag{27}$$

that is, if the error term $\tau_t \neq 0$, then the benchmarking constraint for year t is non-binding.[11] This is a more general setting that considers the benchmarking inconsistencies arising not only by errors in the indicator series, but also caused by errors in the benchmark series.

Finally, by defining

$$\tilde{\mathbf{y}} = \begin{bmatrix} \mathbf{y} \\ \Delta\mathbf{z} \end{bmatrix}, \quad \boldsymbol{\Gamma} = \begin{bmatrix} \mathbf{B} \\ \mathbf{C}_A \end{bmatrix}, \quad \text{and} \quad \mathbf{e} = \begin{bmatrix} \boldsymbol{\tau} \\ \boldsymbol{\varepsilon} \end{bmatrix}, \tag{28}$$

constraints (26) and (27) together form the familiar linear inverse problem with noise formulation $\tilde{\mathbf{y}} = \boldsymbol{\Gamma}\mathbf{x} + \mathbf{e}$ given in (16). Hence, we expressed the AFD-based benchmarking problem in the regression-type equation (16).

The question is now whether we can estimate $\mathbf{x}$ using the traditional econometric methods such as OLS? We think that running OLS on (16) is meaningless from the following practical considerations. First, note that in this regression there are N parameters and $T + N - 1$ observations, implying that the number of degrees of freedom is $df = (T + N - 1) - N = T - 1$. Given that the revision policies of many national statistical agencies restrict the number of years to be revised during the annual revisions to merely 3 years (see Section 1), OLS regression will definitely

[11] In an enhanced version of the Denton procedures non-binding constraints are accommodated by putting the annual constraints in the objective function of an unconstrained minimization problem with a user-specified annual weights. These weights specify how binding are treated the annual benchmarks (see e.g., Bloem et al. 2001, p. 101).

suffer from the problem of insufficient degrees of freedom. Second, in practical applications often the benchmark series are considered reliable, or equivalently in (28) it should be the case that $\boldsymbol{\tau} = \mathbf{0}$. However, guaranteeing that some or all of the benchmarking constraints are binding is impossible in the OLS setting. Our suggestion is to use the GCE approach, discussed in Section 4.1, in tackling this sort of "ill-posed" problem. Namely, we define the prior distributions of $\mathbf{x}$ and $\mathbf{e}$ such that their corresponding prior means are $\mathbf{Rq} = \mathbf{z}$ and $\mathbf{Vu} = \mathbf{0}$, respectively. If any (or all) benchmarking constraint binds, this will be explicitly introduced in the related prior distribution(s) (the details are discussed in Section 5 below). Note that our choice of the prior means is not arbitrary. The MDI principle underlying the GCE approach makes sure that the posterior distributions of $\mathbf{x}$ and $\mathbf{e}$ being consistent with the model constraints (16) are also as close as possible to their prior distributions. This means that the derived benchmarked series is close to the indicator series, but at the same time the involved errors of non-binding constraints are close to zero either. The last outcome reflects the AFD principle behind our entropy-based benchmarking method.

4.3 PFD entropy-based benchmarking method

In this section we develop the alternative entropy-based benchmarking method that is based on the proportional first difference (PFD) preservation principle. This type of movement preservation requires that period-to-period proportional changes of the indicator series are maximally kept in the benchmarked series. Recall that this is the core of the PFD Denton procedure as follows from its objective function (5). Instead of minimization of the overall sum of squares of these proportional changes, we try to maximally implement the PFD principle by having as small as possible error terms ε_j^0 in

$$\frac{x_j}{z_j} = \frac{x_{j-1}}{z_{j-1}} + \varepsilon_j^0 \quad \text{for all } j = 2, \dots N. \tag{29}$$

Growth rates $g_j = z_j/z_{j-1}$ for all $j = 2, \dots, N$ are observable from the preliminary series. Define a new error term as $\varepsilon_j = z_j \varepsilon_j^0$ and a PFD constraint matrix as

$$\mathbf{C}_P = \begin{bmatrix} g_2 & -1 & 0 & \cdots & 0 \\ 0 & g_3 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & g_N & -1 \end{bmatrix}.$$

Then, premultiplication of (29) by z_j yields $0 = -x_j + g_j x_{j-1} + \varepsilon_j$ for all $j = 2, \ldots, N$, hence the PFD constraints (29) can be compactly written as

$$\mathbf{0} = \mathbf{C}_P \mathbf{x} + \boldsymbol{\varepsilon}. \tag{30}$$

Finally, for the PFD entropy-based benchmarking approach, two of the matrices in (28) are redefined as

$$\tilde{\mathbf{y}} = \begin{bmatrix} \mathbf{y} \\ \mathbf{0} \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Gamma} = \begin{bmatrix} \mathbf{B} \\ \mathbf{C}_P \end{bmatrix}. \tag{31}$$

Therefore, constraints (27) and (30) are again written in the form of the linear inverse problem with noise formulation (16). The vectors of unknown are again recovered using the GCE approach by defining prior distributions for $\mathbf{x}$ and $\mathbf{e}$ with the same means as in the case of the AFD entropy-based method. Final discussions in Section 4.2 are also valid for the PFD entropy-based benchmarking.

By now it should be clear that within the GCE framework the main differences in the AFD and PFD preservation principles that underly the discussed entropy-based benchmarking methods technically show up in the structure (content) of the vector of observables $\tilde{\mathbf{y}}$ and the design matrix $\boldsymbol{\Gamma}$. Further, the fact that the error terms in the PFD constraints are $\varepsilon_j = z_j \varepsilon_j^0$, the sizes of z_j's can be useful in defining the scales of extreme values of these errors' prior distributions.

5 Illustrative examples

In this section we evaluate the performance of the (generalized) pro rata distribution, modified additive first difference (AFD) and proportional first difference (PFD) variants of the Denton method, Causey-Trager (CT) growth rate preservation approach, and the AFD and PFD entropy-based benchmarking methods. For this purpose, we examine two illustrative examples. In the first illustration we use Denton's series as used in his 1971 classic paper on benchmarking. For the second example, we consider the time series of changes in inventories of the Netherlands.

The preliminary series in Denton (1971) are $\mathbf{z}_t = (50\ 100\ 150\ 100)'$ for $t = 1, \ldots, 5$, while the benchmark series of annual totals is $\mathbf{y} = (500\ 400\ 300\ 400\ 500)'$. Note that the annual totals of the original series are all equal to 400, thus are inconsistent with the benchmark data $\mathbf{y}$. To give a quantitative flavor to our evaluation test, we compute the following aggregate indicators of closeness of the benchmarked

series $\mathbf{x}$ derived using the above-mentioned methods to the indicator series:

1. Average absolute level difference $AALD = \big(\sum_{j=1}^{N} |x_j - z_j|\big)/N$,
2. average absolute change difference (Chen 2007)

$$AACD = \frac{\sum_{j=2}^{N} |(x_j - x_{j-1}) - (z_j - z_{j-1})|}{N-1} = \frac{\sum_{j=2}^{N} |(x_j - z_j) - (x_{j-1} - z_{j-1})|}{N-1},$$

3. average absolute proportional difference (see e.g., Titova et al. 2010)

$$AAPD = \frac{100}{N-1} \sum_{j=2}^{N} \Big| \frac{x_j}{x_{j-1}} - \frac{z_j}{z_{j-1}} \Big|,$$

4. average absolute benchmark-to-indicator (BI) difference

$$AABID = \frac{100}{N-1} \sum_{j=2}^{N} \Big| \frac{x_j}{z_j} - \frac{x_{j-1}}{z_{j-1}} \Big|,$$

5. average absolute relative proportional difference (Chen 2007)

$$AARPD = \frac{100}{N-1} \sum_{j=2}^{N} \Big| \left(\frac{x_j}{x_{j-1}} - \frac{z_j}{z_{j-1}} \right) / \frac{z_j}{z_{j-1}} \Big|,$$

which is totally equivalent to the average absolute relative BI difference

$$AARBID = \frac{100}{N-1} \sum_{j=2}^{N} \Big| \left(\frac{x_j}{z_j} - \frac{x_{j-1}}{z_{j-1}} \right) / \frac{x_{j-1}}{z_{j-1}} \Big|.$$

The performance indicators of the benchmarked series derived by the above-mentioned methods are given in Table 1. Among the already discussed indicators we also provide the corresponding values for the claimed "ideal" objective function of the CT approach (9) and the degree of smoothness of the BI ratios series, *Smooth*, to be discussed below.

The required supports and prior distributions for the entropy-based methods are chosen as follows. The support space of the benchmarked series for both the AFD and PFD entropy-based benchmarking method (entropy, for short) is taken as $\mathbf{r}_j = (0.1\ 0.55\ 1\ 1.45\ 1.9)' \times z_j$ with the uniform prior $\mathbf{q}_j = (0.2\ 0.2\ 0.2\ 0.2\ 0.2)'$ for all $j = 1, \ldots, 20$. This implies that the prior mean of the unknown vector $\mathbf{x}$ is equal to the original series, $\mathbf{Rq} = \mathbf{z}$, however x_j's are given equal chance of taking any value in their support in the adjustment process. The support space of the disturbance

$\mathbf{e}$ in (16) for the AFD Entropy and PFD Entropy1 is $\mathbf{v}_i = (-3\ 0\ 3)' \times 20$ for all $i = 1, \ldots, 24$.[12] Given the structure of the disturbance term in (28) and the fact that we want the benchmarking constraints (27) to be binding (i.e., $\boldsymbol{\tau} = \mathbf{0}$), we choose the prior $\mathbf{u}_i = (0\ 1\ 0)'$ for errors of the benchmarking constraints, τ_i's, and $\mathbf{u}_i = (0.1\ 0.8\ 0.1)'$ for the AFD/PFD constraints errors, ε_i's. The last prior indicates that we want the disturbance posterior mean $\boldsymbol{\varepsilon}$ to be close to zero so that the derived benchmarked series satisfies to a sufficient degree the AFD and PFD principles discussed in Sections 4.2 and 4.3. PFD Entropy2 in Table 1 has a different disturbance support than that of the PFD Entropy1. Namely, the support of ε_j is chosen as $\mathbf{v}_j = 0.1 \times (-z_j\ 0\ z_j)'$ making use of the additional fact that in Section 4.3 we have defined the corresponding error as $\varepsilon_j = z_j \varepsilon_j^0$ for all $j = 2, \ldots, N$. Hence, in this case the extreme values of the disturbances support are varying and proportional to the size of the corresponding indicator series. We should mention that our results are generally robust to different choices of the number of outcomes in the supports spaces and various size parameters defined above.

Table 1: Aggregate performance indicators: Denton's data

Method	f_{CT}		*AALD*		*AACD*		*AAPD*		*AABID*		*AARPD*		*Smooth*	
Pro rata distribution	0.07	2	15.00	1	9.87	3	2.72	1	5.26	1	5.44	1	6.57	7
AFD Denton	1.20	7	18.32	7	5.97	1	17.51	7	13.07	7	13.26	7	6.41	6
AFD Entropy	0.40	6	17.36	4	6.83	2	10.96	6	9.18	6	9.23	6	4.26	4
PFD Denton	0.14	4	17.56	5	10.56	6	6.97	5	5.94	5	6.09	5	2.48	1
PFD Entropy1	0.07	3	15.77	2	10.41	5	4.82	3	5.68	3	5.84	3	4.91	5
PFD Entropy2	0.15	5	17.56	6	10.64	7	6.97	4	5.92	4	6.07	4	2.49	2
CT (PFD Denton)	0.04	1	16.55	3	10.35	4	3.76	2	5.67	2	5.76	2	3.81	3

Notes: The first three indicators are expressed in absolute terms, while the rest are given in percentages. CT (PFD Denton) means that CT solution is based on having the PFD Denton solution as a starting point in its derivation. Numbers to the right of indicators values show the ranking of each method according to those indicators.

As expected, Table 1 shows that the CT solution results in the lowest value of the sum of squares of the period-to-period growth rates differences of the benchmarked and indicator series, f_{CT}. Following the literature, we used the PFD Denton solution as a starting value in its iterative algorithm derivation. However, we should note that since the CT problem is highly nonlinear, a different starting point might very well end up in a quite different solution possibly with a smaller value of its

[12] The error support is in fact rather wide. Usually, it is defined using the well-known 3σ rule, where σ refers to the standard deviation of the disturbance. According to Chebyshev's inequality, for *any* random variable x with mean μ and finite variance σ^2, the proportion of the distribution's values within v standard deviations from the mean is *at least* $1 - 1/v^2$, that is, in probability terms, $\Pr(|x - \mu| < v\sigma) \geq 1 - 1/v^2$. In case of 3σ rule, one chooses $v = 3$, hence it excludes *at most* one-ninth (or 11%) of any distribution values. For normal distributions, the proportion excluded by the 3σ rule is only 0.27%. Hence, essentially in our case we have assumed that $\sigma = 20$.

objective function. For Denton's example we also run the CT problem with the pro rata distribution and AFD Denton solutions as initial values, but the results were no different from that based of the PFD Denton solution. We also observe that pro rata distribution and PFD Entropy1 take, respectively, the second and third rankings in terms of the growth rate preservation indicator, f_{CT}.

AFD Denton solution outperforms the rest on the base of the $AACD$ indicator, which is again not surprising given the close similarity of this indicator to the modified AFD Denton objective function – the second term in (4). Observe that the majority of the performance indicators discussed above miss a very crucial point of the benchmarking problem. Table 1 shows that the pro rata distribution is doing quite well according to all first six indicators, and even takes the first positions in four cases. The fact that it scores highest on the $AALD$ indicator is totally expectable since the pro rata distribution is all about keeping the levels of original elements in the corresponding benchmarked series (see Section 3). However, as we pointed out in Section 3, in general, the (generalized) pro rata distribution is not an acceptable benchmarking method due to its resulting step problem. That is, it results in discontinuities in the derived time series because any discrepancy in the annual values is put into the corresponding single quarter, whose adjustment is totally independent of other quarterly series. That is why Bloem et al. (2001) advocate using a benchmark-to-indicator (BI) ratio framework in benchmarking. The step problem can be easily recognized from the charts of the BI ratios, which for our example are given in Figure 1.

Figure 1 shows that the curve of the BI ratios of the pro rata method is constant for each year and makes downward and upward abrupt jumps from year to year. This is what exactly benchmarking experts mostly wish to avoid and try to have a benchmarked series with as smoothed as possible BI ratios curve. In order to take the step problem into account, we introduce a *smoothness* indicator, which is defined as

$$Smooth = \frac{100}{N}\sum_{j=1}^{N}|h_j - h_j^s|,$$

where $h_j = x_j/z_j$ is the BI ratio and h_j^s is the corresponding value of the smoothed series of h_j's using a 7-point moving average filter.[13]

Using the smoothness indicator we indeed find that the pro rata distribution produces the worst benchmarked series. Note that according to this measure, PFD

[13]The first elements of the smoothed series are $h_1^s = h_1$, $h_2^s = (\sum_{j=1}^{3} h_j)/3$, $h_3^s = (\sum_{j=1}^{5} h_j)/5$, $h_4^s = (\sum_{j=1}^{7} h_j)/7$, $h_5^s = (\sum_{j=2}^{8} h_j)/7$, etc.

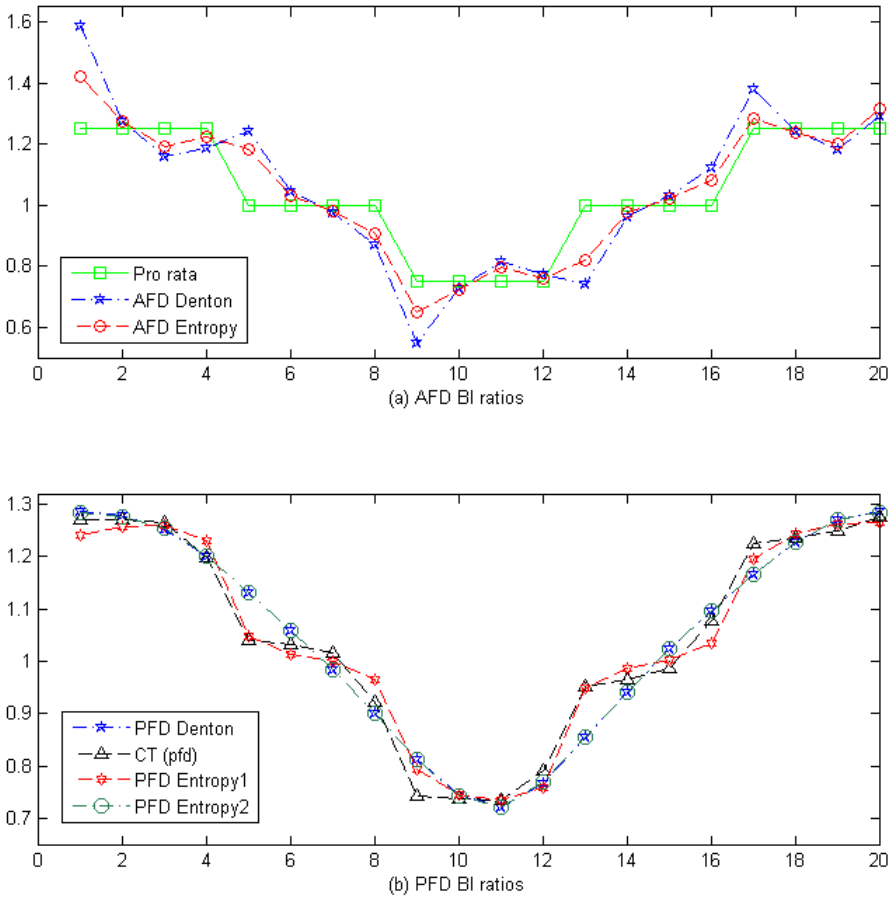


Figure 1: Benchmark-to-indicator ratios for Denton's data

Denton outperforms all the other methods. This is clearly observed by its smoothed BI ratio curve given in chart (b) of Figure 1. Note, however, that according to this measure PFD Entropy2 practically is equivalent to the PFD Denton benchmarked series. This is also confirmed in Figure 1 where the BI ratios curves of these two methods coincide. The CT's growth preservation solution comes third in this ranking and has much less smooth BI ratios curve than those of the PFD Denton and PFD Entropy2 approaches.

Finally, from Table 1 we observe that the AFD Entropy solution outperforms its counterpart AFD Denton results according to all indicators, except $AAPD$. Similarly, the PFD entropy-based outcomes, on average, do as well as the PFD Denton estimates.

Our second example focuses on a volatile time series that allow for both positive

Table 2: Changes in inventories, 2003-2005, The Netherlands (2005 prices, mln euro)

Period	True **x**	**z**	Pro	AFD Den.	PFD Den.	CT(pro)	CT(afd)	CT(pfd)	AFD Ent.	PFD Ent.
2003.Q1	-159	-159	-268	-517	-77	-439	1220900	-53091	-288	-225
2003.Q2	-1128	-752	-1269	-1082	-311	-2074	6110641	-251085	-1220	-1097
2003.Q3	51	128	76	-147	2	172	-14382399	29968	23	60
2003.Q4	467	1168	692	976	-383	1572	7050089	273439	716	493
2004.Q1	132	132	132	51	-21	138	422272	30539	83	128
2004.Q2	-1109	-1109	-1109	-1113	13	-1157	-3516447	-256605	-1108	-1079
2004.Q3	552	552	552	590	174	565	1062033	67179	575	544
2004.Q4	1323	1323	1323	1370	732	1353	2033040	159785	1348	1305
2005.Q1	-10	-10	-10	12	-6	-10	-42198	1075	-10	-10
2005.Q2	-1167	-1167	-1167	-1164	-698	-1148	-4926550	125443	-1167	-1167
2005.Q3	-16	-16	-16	-25	-11	-16	-44786	1135	-16	-16
2005.Q4	1791	1791	1791	1775	1313	1772	5014133	-127055	1791	1791
Performance indicators of the benchmarked series with respect to the true series										
Sign-change cases		0	0	2	3	0	3	4	0	0
f_{CT}		5.6	5.6	47897	43510	5.6	103	6	514.3	5.8
$AALD$		96.2	41.7	111	417	215.5	3818355	114538	49.6	15.9
$AACD$		181.8	56.6	194	736	366.7	7256713	200067	70.3	24.1
$AAPD$		24.5	22.8	2642	2273	24.0	137	27	275.4	29.5
$AABID$		33.5	12.8	174	26	44.1	5396581	10982	34.6	8.0
$AARBID$		33.7	9.0	126	489	17.5	508	48	42.3	6.7
$Smooth$		32.0	8.6	81	14	40.1	3838881	8242	13.4	3.5

Notes: The true series are available from the Statistics Netherlands (www.cbs.nl). The abbreviations are: Pro – generalized pro rata distribution, AFD Den. – AFD Denton method, CT(pro) – CT method with the pro rata solution taken as its initial point, PFD Ent. – PFD entropy-based method. **x** and **z** are the benchmarked and indicator series, respectively.

and negative values. We choose quarterly series of changes in inventories for the Netherlands covering the period from 2003 to 2005. The choice of the number of years is not arbitrary and is consistent with some national revision policies that allow only for 3 years of monthly/quarterly series to be revised during the annual revisions. We then made adjustments to only the last three values of the first year in the true series, and consider them as our indicator series. Both the true and indicator series along with the estimates of the different methods and the performance indicators values of the derived series with respect to the true series are given in Table 2.

Note that we have reduced the second element of the first quarter by a factor of 1.5 and increased the consequent two entries by a factor of 2.5. As a result the sum of the first quarter in the indicator series becomes 385, while the corresponding benchmark value is -769. Such switches in signs of annual data can occur very often with volatile series. However, we should reiterate that this is an illustrative example, whose purpose is simply to see how sensitive the outcomes of the benchmarking methods are to the volatile data that allow for both positive and negative values.

For our AFD and PFD entropy-based approach we choose similar supports and uniform prior as before. That is, the support of the benchmarked series is chosen as $\mathbf{r}_j = (0.1\ 0.55\ 1\ 1.45\ 1.9)' \times z_j$ if $z_j > 0$ and $\mathbf{r}_j = (1.90\ 1.45\ 1\ 0.55\ 0.1)' \times z_j$ if $z_j < 0$. The support space of the disturbance for both the AFD Entropy and PFD Entropy is taken as $\mathbf{v}_i = (-3\ 0\ 3)' \times 150$ with similar prior distribution as before.

Without going into the details, the main point derived from Table 2 is that Denton's methods and CT approach may violate the principle of movement and

sign preservation. We see 2 to 4 sign-changing cases in each benchmarked series of these methods.[14] Also note that the CT solutions based on the AFD and PFD Denton outcomes (as initial point in their derivation) are quite sensitive scale-wise to the series signs volatility. The derived CT solutions are very large in absolute value compared to the indicator series, thus are useless from a practical point of view. This outcome maybe expected due to the high nonlinearity of the CT problem, whose solution becomes unstable in cases of sign-volatile series. The bad performance of the two CT solutions maybe explained (partly) by the fact that the choice of the AFD and PFD Denton solutions as initial points is already questionable as the last suffer from the sign-switching problem themselves. Note, however, that the CT solution using the generalized pro rata series as its initial point provides rather satisfying results. Hence, the second point based on these experiments worth mentioning is that in case of sign-volatile series it maybe preferable to use the generalized pro rata distribution estimates as the initial points in the derivation of the CT solution, because they by construction do not change the signs of the indicator series.

Finally, note that the entropy-based benchmarking methods provide satisfying results, and clearly outperform their AFD and PFD Denton counterparts. The entropy-based benchmarked series automatically satisfy the principle of movement and sign preservation as its requirements are explicitly included into the structure of the relevant support spaces and prior distributions for the unknown variables of the benchmarking problem within the GCE estimator (20)-(23). From Table 2, for example, it follows that PFD Entropy, on average, outperforms all other methods if we consider the performance indicators of the benchmarked series relative to the true series.

6 Conclusion

In this paper we considered alternative benchmarking methods that are based on a cross-entropy formalism. Given that one might deal with volatile time series that allow for both positive and negative values, we argue that such benchmarking should be based on the principle of movement and sign preservation. This principle

[14]The change in sign is a potential problem of using methods with quadratic or higher order objective functions. One way to deal with this issue is to nullify sign-changing elements in the optimization process by introducing an appropriate penalty function in the objective used. See, for example, Temurshoev et al. (2011). Otherwise, one can simply use the non-negativity constraints. In any case the Denton variants then lose their attractiveness as they will no longer have analytical solutions and numeral methods are called for.

states that (i) short-term movements in the observed high-frequency data should be maximally reproduced in the benchmarked series and (ii) the signs of the original elements should be kept in the corresponding benchmarked series. Earlier literature, however, focuses only on the first part of this principle, i.e., it regards the principle of movement preservation as the main driver of any acceptable benchmarking approach.

We show that the widely used additive first difference (AFD) and proportional first difference (PFD) variants of Denton's (1971) adjustment procedure and Causey and Trager's (1981) growth rate preservation method, in general, do not satisfy the above-mentioned principle. They can produce benchmarked series with switching signs compared to their indicator series in case of benchmarking of sign-volatile series. Further, we generalize the so-called pro rata distribution method to benchmarking of sign-volatile time series, hence provide theoretical foundation for this level preservation method.

Our proposed AFD and PFD entropy-based benchmarking methods make use of a generalized cross-entropy (GCE) estimator of Golan et al. (1996), thus satisfy the principle of movement and sign preservation by construction. This is because the prior beliefs about the signs, magnitudes, and/or ranges of plausible values of the unknown variables can be easily incorporated into the corresponding GCE problem. The AFD and PFD entropy-based benchmarking methods have several advantages. They

- allow for any choice of binding and/or nonbinding benchmarking constraints,
- allow using explicitly crucial non-sample information regarding the signs, size and/or ranges of reasonable values of benchmarked elements,
- are applicable for any size of yearly observations, and
- provide reliability factors for each element of the benchmarked series.

One might argue that the entropy-based outcome is subjective and depends on the supports and priors choices. We, however, consider this as the entropy approach flexibility. That is, experts' crucial non-sample information can be taken into full account by appropriate choice of the supports and priors that would improve the final estimates.

Our illustrative examples confirm that the AFD and PFD Denton methods and Causey-Trager (CT) solution (based on the AFD/PFD Denton solution taken as its initial point) may suffer from the sign-switching problem. In fact, due to this possibility, the CT solution may become unstable and practically useless. Our results show that in such cases using the generalized pro rata distribution benchmarked

series as initial point in the derivation of the CT solution may very well avoid this instability issue. Finally, the illustrative example demonstrates that the proposed AFD and PFD entropy-based benchmarking methods maybe considered plausible competitors to current benchmarking methods, and maybe even preferred in some circumstance, i.e., when the time series of interest is sign-volatile.

Appendix

Derivation of the generalized pro rata distribution formula

In what follows we apply the well-known RAS updating idea to the benchmarking problem.[15] For simplicity purposes, we make the distinction between annual and sub-annual data explicit. That is, for example, instead of z_j in the derivation it is more convenient to write z_{it} to refer to the indicator value of sub-period $i = 1, \ldots, M$ for year $t = 1, \ldots, T$. Using a mathematical trick used by Junius and Oosterhaven (2003) in setting a generalized RAS framework for preserving the signs of the original elements in the estimated ones, we define $h_{it} \equiv x_{it}/z_{it}$ for all $i = 1, \ldots, M$ and all $t = 1, \ldots, T$ whenever $z_{it} \neq 0$, and set $h_{it} = 1$ for $z_{it} = 0$.

The benchmarking constraints $\boldsymbol{\imath}'\mathbf{x}_t = y_t$ in terms of h_{it}'s can be easily written as

$$\sum_{i=1}^{M} z_{it} h_{it} = y_t \qquad \text{for all } t = 1, \ldots, T. \tag{32}$$

Next, let us define $\mathbf{P}$ as a matrix with all non-negative entries of $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_T)$, and $\mathbf{N}$ as a matrix containing absolute values of the negative elements of $\mathbf{Z}$, i.e., $\mathbf{Z} = \mathbf{P} - \mathbf{N}$. Hence, using the objective (12) the associated Lagrangean of the GRAS-like benchmarking problem is

$$\mathcal{L} = \sum_{(i,t)\in\mathbf{P}} z_{it}\left(h_{it}\ln\left(\frac{h_{it}}{e}\right)+1\right) - \sum_{(i,t)\in\mathbf{N}} z_{it}\left(h_{it}\ln\left(\frac{h_{it}}{e}\right)+1\right) + \sum_{t=1}^{T}\lambda_t\Big(y_t - \sum_{i=1}^{M} z_{it}h_{it}\Big),$$

where λ_t is the Lagrange multiplier of the annual constraints (32). The optimal solutions of this function can be easily derived as

$$h_{it} = \begin{cases} e^{\lambda_t} & \text{if } z_{it} \geq 0, \\ e^{-\lambda_t} & \text{if } z_{it} < 0. \end{cases} \tag{33}$$

Note that for the solution of minimizing (12) with respect to (32) it always holds that $h_{it} > 0$, which means that the estimated values will preserve the signs of the original elements. For simplicity, denote $s_t \equiv e^{\lambda_t}$. Then, using $\mathbf{z}_t = \mathbf{p}_t - \mathbf{n}_t$ and the definition of h_{it}, the optimal solution (33) becomes $\mathbf{x}_t = \mathbf{p}_t \times s_t - \mathbf{n}_t/s_t$, which is given in (14).

The annual adjustment factors are derived using the benchmarking constraints (32). Using (14) in the annual constraint $\boldsymbol{\imath}'\mathbf{x}_t = y_t$ gives $\boldsymbol{\imath}'\mathbf{p}_t \times s_t - \boldsymbol{\imath}'\mathbf{n}_t/s_t = y_t$. Premultiplying the last equation by s_t yields $\boldsymbol{\imath}'\mathbf{p}_t \times s_t^2 - y_t s_t - \boldsymbol{\imath}'\mathbf{n}_t = 0$, which is a quadratic equation in s_t that admits two solutions. However, for our purposes we need only its positive root,

[15] See Lahr and de Mesnard (2004) for details on RAS (including its history), which also gives an extensive set of references on the topic.

thus $s_t = 0.5 \times \left[y_t + \sqrt{y_t^2 + 4 \times (\boldsymbol{\iota}'\mathbf{p}_t) \times (\boldsymbol{\iota}'\mathbf{n}_t)}\right] / \boldsymbol{\iota}'\mathbf{p}_t$ whenever $\boldsymbol{\iota}'\mathbf{p}_t > 0$. If, however, $\mathbf{p}_t = \mathbf{0}$, then $\mathbf{n}_t = -\mathbf{z}_t$. Consequently, the mentioned quadratic equation boils down to $\boldsymbol{\iota}'\mathbf{0}s_t^2 - y_t s_t - \boldsymbol{\iota}'\mathbf{n}_t = 0$. Hence, $s_t = -\boldsymbol{\iota}'\mathbf{n}_t / y_t = \boldsymbol{\iota}'\mathbf{z}_t / y_t$ if $\mathbf{p}_t = \mathbf{0}$. These two cases of the annual adjustment factors are given in (15).

Can one apply the pro rata distribution to benchmarking stock variables? In such cases the $M \times 1$ coefficient vector of annual constraints takes the form $\mathbf{b} = (\mathbf{0}', \ 1)'$. If one uses the corresponding annual constraints of $z_{Mt} h_{Mt} = y_t$ for all $t = 1, \ldots, T$ instead of (32), the result is completely unsatisfying: adjustments will be made only on the last sub-annual entries, i.e., on z_{Mt} to obtain new x_{Mt} for all t, while the rest will remain equal to the original data, i.e., $x_{it} = z_{it}$ for all $i = 1, \ldots, M-1$ and all t. From a practical point of view, however, this outcome is not at all useful. More useful adjustment would be spreading the difference between the last sub-annual value of the stock variable and the corresponding annual value over all the corresponding sub-annual elements. That is, instead of the cumulative values of the high-frequency data, one could use their *marginal changes* in the indicator series. As a result one will again end up with the annual constraints (32) that are applicable for flow variables. In the final stage then the derived marginal values can be easily transformed into the cumulative values which will be the benchmarked series of the stock variable of interest.

References

Bikker, R., J. Daalmans, and N. Mushkudiani: 2010, 'A multivariate Denton method for benchmarking large data sets'. *CBS Discussion paper* **10002**.

Bloem, A. M., R. J. Dippelsman, and N. Ø. Mæhle: 2001, *Quarterly National Accounts Manual: Concepts, Data Sources, and Compilation*. Washington D.C.: International Monetary Fund.

Boot, J. C. G., W. Feibes, and J. H. C. Lisman: 1967, 'Further methods of derivation of quarterly figures from annual data'. *Applied Statistics* **16**, 65–75.

Brown, I.: 2010, 'An empirical comparison of constrained optimization methods for benchmarking economic time series'. *Business and Economic Statistics Section - JSM 2010* pp. 2131–2143.

Causey, B. and M. L. Trager: 1981, 'Derivation of solution to the benchmarking problem: trend revision'. Unpublished research note, U.S. Census Bureau.

Chen, B.: 2007, 'An empirical comparison of methods for temporal disribution and interpolation in the national accounts'. BEA working paper WP2007-03.

Chen, B. and S. H. Andrews: 2008, 'Research spotlight: an empirical review of methods for temporal disribution and interpolation in the national accounts'. *Survey of Current Business* **88**, 31–37.

Chow, G. C. and A.-l. Lin: 1971, 'Best linear unbiased interplation, distribution and extrapolation of time series by related series'. *Review of Economics and Statistics* **53**, 372–375.

Cover, T. M. and J. A. Thomas: 2006, *Elements of Information Theory*. New Jersey: John Wiley and Sons, 2 edition.

Dagum, E. B. and P. A. Cholette: 2006, *Benchmarking, Temporal Distribution, and Reconciliation Methods for Time Series*. New York: Lecture Notes in Statistics 186, Springer.

Denton, F. T.: 1971, 'Adjustment of monthly or quarterly series to annual totals: an approach based on quadratic minimization'. *Journal of the American Statistical Association* **66**, 99–102.

Di Fonzo, T. and M. Marini: 2010, 'Benchmarking and movement preservation. Evidences from real-life and simulated series'. University of Padua, Working Paper 14.

Eurostat: 1999, *Handbook on Quarterly National Accounts*. Luxembourg: Office for Official Publications of the European Communities.

Golan, A., G. Judge, and D. Miller: 1996, *Maximum Entropy Econometrics: Robust Estimation with Limited Data*. Chichester, UK: John Wiley and Sons.

Helfand, S. D., N. J. Monsour, and M. L. Trager: 1977, 'Historical revision of current business survey estimates'. *Proceedings of the Business and Economics Statistics Section, American Statistical Association* pp. 246–250.

Janes, E. T.: 1957a, 'Information theory and statistical mechanics'. *Pysical Review* **106**, 620–630.

Janes, E. T.: 1957b, 'Information theory and statistical mechanics II'. *Pysical Review* **108**, 171–190.

Junius, T. and J. Oosterhaven: 2003, 'The solution of updating or regionalizing a matrix with both positive and negative entries'. *Economic Systems Research* **15**, 87–96.

Kullback, S.: 1959, *Information Theory and Statistics*. New York: John Wiley and Sons.

Kullback, S. and R. A. Leibler: 1951, 'On information and sufficiency'. *Annals of Mathematical Statistics* **22**, 79–86.

Lahr, M. L. and L. de Mesnard: 2004, 'Biproportional techniques in input-output analysis: table updating and structural analysis'. *Economic Systems Research* **16**, 115–134.

Lenzen, M., R. Wood, and B. Gallego: 2007, 'Some comments on the GRAS method'. *Economic Systems Research* **19**, 461–465.

Shannon, C. E.: 1948, 'A mathematical theory of communication'. *Bell System Technical Journal* **27**, 379–423.

Temurshoev, U. and M. P. Timmer: 2011, 'Joint estimation of supply and use tables'. *Papers in Regional Science* **90**, 863–882.

Temurshoev, U., C. Webb, and N. Yamano: 2011, 'Projection of supply and use tables: methods and their empirical assessment'. *Economic Systems Research* **23**, 91–123.

Titova, N., D. F. Findley, and B. C. Monsell: 2010, 'Comparing the Causey-Trager method to the multiplicative Cholette-Dagum regression-based method of benchmarking sub-annual data to annual benchmarks'. *Business and Economic Statistics Section - JSM 2010* pp. 3007–3021.